AD-A250 344

ENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)

2. REPORT DATE
1989

3. REPORT TYPE AND DATES COVERED
Unknown

4. TITLE AND SUBTITLE
Against Conditional Probability

5. FUNDING NUMBERS
DAAB10-86-C-0567

6. AUTHOR(S)
Henry E. Kyburg, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Rochester
Department of Philosophy
Rochester, NY 14627

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army CECOM Signals Warfare Directorate
Vint Hill Farms Station
Warrenton, VA 22186-5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
92-TRF-0033

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Statement A; Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
While a minority of statisticians hew to the Bayesian line, a large number of philosophers and a large number of AI researchers take Bayesian conditionalization for granted as the only way of updating uncertainties. At the same time, everybody, Bayesian or not, appears to accept the fundamental principle of direct inference: if you know the statistics, the statistics should constrain your belief. The contribution of this paper is to exhibit a conflict between these two principles, and to argue in favor of direct inference and against conditionalization.

14. SUBJECT TERMS
Conditional Probabiltiy, Artificial Intelligence, Data Fusion
Uncertain Inference, Inverse Probabiltiy

15. NUMBER OF PAGES
15

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

AGAINST CONDITIONAL PROBABILITY

by

Henry E Kyburg, Jr.

Department of Philosophy

University of Rochester

Rochester N Y 14627

kyburg@rochester arpa

Automated Reasoning

Uncertainty, Evidential Updating

Science

While a minority of statisticians hew to the Bayesian line, a large number of philosophers and a large number of AI researchers take Bayesian conditionalization for granted as the only way of updating uncertainties. At the same time, everybody, Bayesian or not, appears to accept the fundamental principle of direct inference: if you know the statistics, the statistics should constrain your belief. The contribution of this paper is to exhibit a conflict between these two principles, and to argue in favor of direct inference and against conditionalization.

Against Conditional Probability

1. Background.

If there is any distinction in the realm of statistics or inductive logic or the manipulation of uncertainty that deserves to be called "classical", it is the distinction between direct and inverse inference.

Direct inference includes among its premises some statement of statistical distribution, or relative frequency, or chance, and concludes with a statement of probability concerning a sample or a single case. From 50% of coin-tosses yield heads, in the absence of countervailing arguments, we conclude that the probability is 0.5 that the next toss will yield heads.

Inverse inference takes as its premises a statement of sample statistics concerning a sample from a population, together with some other premises, and concludes with a statement of statistical distribution, or relative frequency, or chance, applicable to the population as a whole.

Both direct and inverse inference are characterized by nonmonotonicity. Adding to the premises may undermine a conclusion in either case. This was recognized explicitly by R. A. Fisher [1936, p. 254]: "There is one peculiarity of uncertain inference which often presents a difficulty to mathematicians trained only in the technique of rigorous deductive argument, namely that our conclusions are arbitrary, and therefore invalid, unless all the data, exhaustively, are taken into account In rigorous deductive reasoning we may make any selection from the data, and any certain conclusions which may be deduced from this selection will be valid, whatever additional data we have at our disposal " Even so, direct inference has been regarded as relatively unproblematic.

Given as a premise that the chance of heads on the toss of a coin is a half, we confidently say that the probability of heads on the first toss is a half Given a further premise to the effect that three of the first four tosses yielded heads, we recompute the probability of heads on the first toss to be 3/4

2. Bayes' Theorem

Inverse inference has often been associated with Bayes theorem. For example, one way of getting at the parameter p characterizing the proportion of black balls in an urn is to draw a sample, and to apply Bayes' theorem Bayes' theorem, however, requires as input a *prior distribution* for the parameter p, which may be difficult to justify in terms of frequencies or chances. Thus Jerzy Neyman, another founding father of modern statistics, writes [Neyman, 1957 p. 7] " . . . persons who would like to deal only with classical probabilities, having their counterparts in the really observable frequencies, are forced to look for a solution of the problem of estimation other than by means of the theorem of Bayes." This is not to say that Bayes' theorem is never applicable. As R. A. Fisher saw clearly, there are many situations in which Bayes' theorem can easily be construed in terms of direct inference. In (Fisher, [1930]) he notes that drawing from a super-population in which the parameter of interest (say p) has a known distribution, and then getting a *posterior* distribution for p, "... is a perfectly direct argument. ..." For inverse inference proper -- that is, inference whose uncertainty is *not* based on known frequencies, but on subjective probabilities -- Fisher has nothing but contempt [1930]. "In fact, the argument runs somewhat as follows: a number of useful but uncertain judgments can be expressed with exactitude in terms of probability; our judgments respecting causes or hypotheses are uncertain, therefore our rational attitude towards them is expressible in terms of probability." Neyman's attitude is even less tolerant.

Fisher and Neyman were reacting against the use of the so-called *axiom of Bayes* that stipulated the use of uniform priors. Their goal, which has informed most of modern statistical practise, was to do without priors. Since their day, however, inverse inference proper has become (almost) respectable again. This is particularly so in philosophy, in which inductive inference is often supposed to take place only by means of Bayes' theorem, and in AI, in which the updating or modification of uncertainty is assumed to take place only by means of conditionalization. I claim that there is a serious conflict between direct inference and inverse inference proper -- that is, the use of conditionalization. Since the original point of inverse inference was to serve the interest of direct inference by providing statistical premises, we should hold onto direct inference and abandon conditionalization except in those cases (which are many) in which it can be reduced to direct inference.

3. Direct Inference.

Direct inference has always seemed so obvious, that amost nobody has made a serious attempt to reduce it to rules. (Reichenbach [1949] is an exception, my contrasting view appeared in [1961].) Two simple rules will suffice for our purposes here, they are the *difference* rule and the *strength* rule. Essentially, they are rules for choosing a reference class.

We will generalize the notion of probability very slightly to accomodate the obvious fact that we don't know most general frequencies or distributions exactly, we will therefore represent both probabilities and our knowledge of frequencies by intervals. We will say that two intervals *differ* if they are not identical and neither is included in the other. If one is included in the other we will say that it is *stronger* than the other.

The *difference* rule says that if you have two possible reference classes, and they are characterized by different frequency intervals, then, if one reference class is included in the other, it may be the right reference class, but if neither is included in the other we may have to look elsewhere for a reference class. Thus if you know of a card that it is black, the probability that it is a spade is determined by the frequency of spades among black cards, not among cards in general.

The *strength* rule says that if you have two possible reference classes, and neither is ruled out as a reference class by differing from some other reference class, then the one about which our statistical knowledge is stronger is the better reference class. As an extreme example, I know that the frequency of heads in the set of tosses consisting of the singleton of the next toss is in the closed interval $[0,1]$, and that's all I know about it. But I know that among coin-tosses in general, heads occur with a frequency very close to a half. It is the latter that constitutes a better reference class.

This characterization of direct inference leaves out a number of important aspects, but they are not essential for our purposes here. Details can be found in [1983].

4. Example.[1]

Here is an example in which the conflict between direct inference and conditionalization comes out clearly

Suppose that we are running a factory, and we use a certain type of instrument to test our product The manufacturer of the instruments certifies on the basis of extensive testing, that his instruments are subject to error greater than e exactly 20% of the time. Of course we understand "exactly 20%" to mean *very close* to 20%. We have no reason to doubt this report.

Pick an item off the assembly line. Test it with the instrument. The probabilty that the true value is within e of the reading is clearly .20. This is just direct inference, making use of the fact that we know the frequency of errors of magnitude e.

Now let us suppose that we seem to notice that some instruments are more accurate than others Of course that is bound to be the case, and does not impugn the manufacturer's claim that the error rate is 20%. We are inspired to look into the matter more deeply: we note that the instruments are inspected by three different inspectors, A, B, and C. We form the hypothesis that the accuracy of the instrument is related to the identity of the inspector who passed it We take a sample of 400 each of readings made on each kind of instrument, and compare the readings made by our super-accurate-tester

The number of readings in error by more than e , and the ratio of such readings to the total number, is presented in the following table

type	trials	errors	rate
A	400	108	270
B	400	52	130
C	400	70	175

We may also compute ratios in broader classes from the same data:

$A \cup B$	800	160	200
$A \cup C$	800	178	2225
$B \cup C$	800	122	1525
$A \cup B \cup C$	1200	230	1917

What do we do with this data? Well, we can use direct inference to draw conclusions about the general classes of measurements A, B, and C, and their combinations. Taking 95 as an acceptance level, we note that that corresponds to $\pm$ 2 standard deviations on the normal distribution. In the present case we can disregard the difference between the binomial and the normal distribution. The mean of the difference between the sample mean and the population mean is 0. We thus get the following confidence intervals for the error rates in the classes tested:

<u>type</u>	<u>number</u>	<u>standard deviation</u>	<u>interval</u>
A	400	040	[230, 310]
B	400	040	[.090, .170]
C	400	.040	[135, 215]
$A \cup B$	800	.028	[.172, 228]
$A \cup C$	800	.028	[.194, .251]
$B \cup C$	800	.028	[124, .181]
$A \cup B \cup C$	1200	.023	[.169, 215]

Assume that the sampling procedures leave nothing to be desired. Note that our results do not impugn the manufacturer's claim that the relative frequency of errors is .20 At the same time, when we now pick an item from the assembly line and measure it, if we notice that we have an instrument of type A, we will suppose that the probability is [.23, .31] that the reading is in error by more than e. Similarly, if we use an instrument of type B, we find the probability of this amount of error to be only [.09, .17].

On the other hand, if we are looking over old records, and the inspector of the instrument with which a measurement was made was not recorded, it seems right to use the old probability of error of .20. Someone *might* want to argue that since we have this new information about $A \cup B \cup C$ -- namely, that the frequency of error is between .169 and .215 -- we should use that interval. But why? That doesn't *disagree* with the manufacturer's error frequency, and no doubt his frequency is based on vastly more information than is ours. It seems to be just a waste of good information to use our rough estimate in place of his refined one. At any rate, this is the intuition on which the strength rule is based.

If these are our intuitions, we must reject conditionalization in this case. Take the probabilities we get from direct inference to constrain our degrees of belief; let our degrees of belief be given by a belief function BF satisfying the axioms of probability. Then by conditionalization and the principle of total probability we have

$$BF(E|A \cup B \cup C) = BF(E|A \cup B)BF(A \cup B) + BF(E|C)BF(C)$$

Since the probability of E given $A \cup B$ is .200, by the strength rule, and the same is true of the probability of E given $A \cup B \cup C$, and since the probability of E|C is positive, this identity can only be satisfied if BF(C) is 0. Again, by conditionalization and the principle of total probability we have

$$BF(E|A \cup B \cup C) = BF(E|A \cup C)BF(A \cup C) + BF(E|B)BF(B)$$

so that, by the same argument, we have BF(B) = 0.

From this it follows that BF(A) = 1, since BF(B) = BF(C) = 0. So we have $BF(E|A \cup B \cup C) = BF(E|A) \in [.230, .310]$, contrary to our assumption that BF was to be constrained by our probability intervals.

We must give up something. The principle of total probability is hard to give up: every frequency function that applies to the world satisfies the principle of total probability. But conditionalization, applied to a belief function, describes how our beliefs are supposed to change in response to incoming evidence. And there is nothing sacrosanct about that. Of course conditionalization will apply sometimes. If we knew what proportion of the measurements were given by each of the three types of instruments (so what we had probabilities on which to base BF(A), BF(B), and BF(C)) then of course we would not apply the strength rule, but rather an appropriate weighted average in obtaining $BF(E|A \cup B \cup C)$. And this could (and should) be based on a direct inference.

5. Discussion

One way of dealing with the problem of the previous section is to insist that once we have tested a number of the measurements made in our factory, we should use those statistics for our probabilities. And thus, for example, to use [.169, .215] rather than .200 for $BF(E|A \cup B \cup C)$. This won't do for two reasons. Suppose we had tested the three types of instruments and *not* found any evidence that the expected frequency of errors differed. Surely in that case we would feel free to continue using the manufacturer's error rate of .200. Furthermore, we always have very specific statistical information concerning the error rate in future measurements. Thus I know that the error rate among future measurements made by me using instruments of type A is in the interval [0,1]. And this is probably *all* I know about that class of measurements. Surely I should not be required to take the probability of error to be [0,1].

The response of the subjectivist to this sort of example is two-fold. First the subjectivist will assert that probabilities *are* belief functions, and that therefore intervals won't do. Given any measurement performed with an instrument in $A \cup B \cup C$, BF(A), BF(B), and BF(C) -- the degree of belief that it was performed with an instrument of type A, B, or C, respectively -- are all real valued and add up to 1. Similarly, the conditional probability $BF(E|A)$ is real valued: by the preceding kind of argument $BF(E|A) = BF(E|A_1 \quad A_2) = \Sigma\, BF(E|A_i) \times BF(A_i)$, where A_i is the condition that an instrument of type A is used *and* the error rate of instruments of type A is in the i'th subinterval of [.230, .310]. Naturally, we can make these subintervals as small as we want. So the subjectivist thinks I can do things that I don't think I can do, like making all probabilities precise.

On the other hand, the subjectivist can offer an independent argument for conditionalization. Since I have accepted total probability, if I can be compelled to accept conditionalization as well, I shall find myself having to reject direct inference -- or at least the strength rule. The argument goes like the standard dutch book arguments for the probability axioms. Roughly it is this: if you allow conditional bets, but do not adjust your beliefs in accordance with the principle of conditionalization, then your unfriendly bettor will bet on X at your odds, on X & Y at your odds, and will make a *conditional* bet on X, conditional on the occurrence of Y, at those odds you will offer once you have observed Y. If you do not obey the principle of conditionalization, these need not be the same odds, and the unfriendly better will be able to win for sure. Thus, it is claimed, the principle of conditionalization has the same degree of soundness as any other principle of probability.

To this I respond that dutch book arguments are not very persuasive anyway. It is a matter of deductive self-preservation -- and has nothing to do with degrees of belief -- not to make a set of bets on which you are bound to lose money. But it is also not always appropriate to look at conditional bets. In the example of the instrument, the numbers to which I am led seem perfectly reasonable, even though they are not consistent with any probabilistic belief function whose range includes prior beliefs about which instrument is used. Finally, as one can see from this kind of example, the subjectivist requires that the value of the belief function be determined for all possible future contingencies -- and then never changed. (Conditionalization does not involve a change of the basic belief function. $P'(E) = P(E \& A)/P(A)$, for example, so that while one's initial absolute belief function is updated, one is never allowed to change one's mind.)

In terms of the book-making metaphor, the bookie must post odds on all possible contingencies, and then take those odds to determine all his conditional bets; whatever happens, whatever new evidence there is, the bookie need merely look up the corresponding conditional odds in his initial table. He cannot change his odds. I suggest that this is overly rigid. One should, perhaps rarely, be willing to change ones odds in a fundamental way. Learning only by conditionalization implies an excessively narrow view of learning. Where there is conflict with direct inference, direct inference should prevail, and conditionalization should go hang.

note:

1. This example is due essentially to Levi [1977] and [1980], where it is alleged to show the incoherence of the strength rule. I have profited also from extensive discussions with Levi on these matters.

bibliography:

Fisher, Ronald A. [1930] "Uncertain Inference," *Proceedings of the American Academy of Arts and Sciences* **71**, 245-254.

Fisher, Ronald A. [1930] "Inverse Probability," *Proceedings of the Cambgidge Philosopphical Society* **26**, 528-535.

Kyburg, Henry E. [1961] *Probability and the Logic of Rational Belief*, Wesleyan University Press, Middletown.

Kyburg, Henry E. [1983] "The Reference Class," *Philosophy of Science* **50**, 374-397.

Levi, Isaac [1977] "Direct Inference," *Journal of Philosophy* **74**, pp 5-29.

Levi, Isaac [1980] *The Enterprise of Knowledge*, MIT Press, Cambridge.

Neyman, Jerzy [1957] "'Inductive Behavior' as a Basic Concept in Philosophy of Science," *Bulletin of the International Statistical Institute* **25**, pp 7–22.